

# Guidelines for Annotating Rule-based Models

Lily A. Chylek<sup>1,2</sup>, Bin Hu<sup>1,2</sup>, and William S. Hlavacek<sup>1,2</sup>

**Short Abstract** — Conventions for illustrating and annotating computational models of biochemical systems are needed for models to be understood and reused. To this end, we have developed the concept of an extended contact map, which represents the scope of a rule-based model by including an illustration of each molecule, molecular component, interaction, and enzyme-substrate relationship accounted for in a model. Elements of the map can be associated with the rules of the underlying model.

**Keywords** — rule-based modeling, contact map, annotation, signal transduction

## I. BACKGROUND

RULE-BASED modeling is a relatively new approach that is well-suited for representing the kinetics of molecular interactions, especially the interactions of molecules composed of modular functional components, such as proteins. A rule can be viewed as a coarse-grained representation of the kinetics of a class of (bio)chemical reactions. Each reaction within a common class involves a common reaction center and transformation. The granularity of a rule can be refined by adjusting the necessary and sufficient conditions that are required of reactants. Thus, a modeler is free to control the coarseness of model assumptions [1,2].

Rules have been used to build detailed models of signal-transduction systems that capture site-specific interactions between proteins [3,4]. An individual rule in a model is easily visualized, but communicating the scope of a model, especially a comprehensive model for a whole pathway, requires the introduction of new conventions for model illustration and documentation. Such conventions are unavailable, even though it is now possible to formulate and simulate models composed of large numbers of rules [5].

## II. RESULTS

We introduce the concept of an extended contact map. An extended contact map complements visualization of individual rules within the rule set of a model. Rules are largely self-annotating, but it is difficult to grasp the full scope of a model when it contains hundreds of rules. The scope of a model is illustrated by an extended contact map, which also provides information that may not be explicitly represented in individual rules but that is important for understanding the intended meaning of rules.

### A. Molecules

Nested boxes are used to show the subunits of complexes

as well as substructures of individual proteins. The locations of residues that undergo post-translational modification are indicated as small squares inside protein boxes. The square is connected to a line segment, where the residue number and the type of modification are specified. If a residue is part of a domain, the square is placed inside the domain box.

### B. Interactions

Rules that share the same reaction center are mapped to an interaction arrow in an extended contact map. The arrows that represent interactions connect the parts of molecules that mediate the interactions. Arrow symbols inspired by Kohn and co-workers [6] are used to show catalysis and non-covalent binding.

### C. Layout

The starting point of a signaling cascade is placed in the top layer of a map. Each successive layer contains molecules further downstream from the initial stimulus. This arrangement is indicative of the sequence of events in signaling, but not of protein locations. To provide information about location, compartment labels are attached to molecule boxes.

### D. Clarification of Modeling Assumptions

An extended contact map aims to represent the biochemical details of the interactions in a system rather than modeling assumptions. For example, multiple residues may be lumped together as a single site in a rule set, but an extended contact map shows each residue individually. Additionally, rules may not explicitly mention the catalytic domain of an enzyme, but this domain is shown in the map to help the reader understand the function of the protein.

## III. CONCLUSION

An extended contact map for a signaling pathway tends to provide a compact and informative representation of a model for the pathway. It can be used to organize the rules of a model as well as to construct a narrative description of the available knowledge upon which the model is based.

## REFERENCES

- [1] Faeder JR, Blinov ML, Hlavacek WS (2009) Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol* **500**, 113-67.
- [2] Hlavacek WS, et al. (2006) Rules for modeling signal-transduction systems. *Sci STKE* **2006**, re6.
- [3] Faeder JR, et al. (2003) Investigation of early events in Fc epsilon RI-mediated signaling using a detailed mathematical model. *J Immunol* **170**, 3769-81.
- [4] Blinov ML, et al. (2006) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems* **83**, 136-151.
- [5] Colvin J, et al. (2009) Simulation of large-scale rule-based models. *Bioinformatics* **25**, 910-7

<sup>1</sup>Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>2</sup>Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA